

PDH: A Human-Centric Interface for Image Libraries

Baback Moghaddam *

Qi Tian †

Neal Lesh *

Chia Shen *

Thomas S. Huang †

TR2002-52 August 2002

Abstract

We present visualization and layout algorithms that can enhance informal story-telling using personal digital data such as photos in a face-to-face social setting. In order to build a more intuitive browser for retrieval, navigation and story-telling, we introduce a novel optimized layout technique for large image sets, which respects (context-sensitive) mutual similarities as visualized on a shared 2-D display (a table-top). The experimental results show a more perceptually intuitive and informative visualization of traditional CBIR-based retrievals, providing not only a better understanding of the query context but also aiding the user in forming new queries. A framework for user-modeling is also introduced and tested. This allows the system to adapt to the user's preference and integrate relevance feedback

*International Conference on Multimedia & Expo (ICME'02)
Lausanne, Switzerland, August 2002*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Copyright © Mitsubishi Electric Research Laboratories, Inc., 2002
201 Broadway, Cambridge, Massachusetts 02139

* MERL - Research Laboratory

† University of Illinois at Urbana-Champaign

Published in: *International Conference on Multimedia & Expo (ICME'02)*, August 2002.

PDH: A Human-Centric Interface for Image Libraries

Baback Moghaddam, Qi Tian, Neal Lesh, Chia Shen, Thomas S. Huang

Mitsubishi Electric Research Laboratory, Cambridge, MA 02139 {baback, lesh, shen}@merl.com
Beckman Institute, University of Illinois, Urbana-Champaign, IL 61801 {qitian, huang}@ifp.uiuc.edu

Abstract

** We present visualization and layout algorithms that can enhance informal storytelling using personal digital data such as photos in a face-to-face social setting. In order to build a more intuitive browser for retrieval, navigation and story-telling, we introduce a novel optimized layout technique for large image sets, which respects (context-sensitive) mutual similarities as visualized on a shared 2-D display (a table-top). The experimental results show a more perceptually intuitive and informative visualization of traditional CBIR-based retrievals, providing not only a better understanding of the query context but also aiding the user in forming new queries. A framework for user-modeling is also introduced and tested. This allows the system to adapt to the user's preference and integrate relevance feedback.*

Introduction

We often find ourselves using photos to "re-connect" with people, whether it is with our family who we have not seen all day, a friend or colleague whom we have not seen in a year, or our parents who live across the country. In this paper, we present visualization and layout algorithms that can enhance informal storytelling using personal digital data such as photos, audio and video in a face-to-face social setting. These algorithms are designed in the context of the Personal Digital Historian (PDH) system, which supports multi-person interactive informal storytelling by combining and extending research in largely two areas: (1) human-computer interaction and interface (the design of the shared-display devices, user interface for story-telling and on-line authoring, and story-listening), (2) content-based information visualization and retrieval (user-guided image layout, data mining and summarization).

2. Motivation and Interface Design

The following are the design principles of the PDH system: (a) the user interface will enable natural face-to-

face conversation – not forcing everyone to face in the same direction or at their own version of the data; (b) the physical display device must be convenient and customary to use – helping to make the computer disappear; (c) the system should enable interactive and exploratory usage scenarios. The initial PDH prototype employs a circular user interface, as shown in Figure 1, with top projection onto a physical table surface.



Figure 1. PDH table

Current research focuses on three specific areas: (1) UI and HCI: User interface on non-rectangular real world objects, easy re-orientation of both the user interface and the content, Cartesian to Polar coordinate visual layout translation algorithms, radial deformation to display large amount of images/contents, multiple input and manipulation points, and natural visual and audio query formulation with minimal menu-driven interaction. (2) Content organization and annotation: Use 4 W's (who, where, when, what) from storytelling as the primary annotation for organization and navigation, an example is shown in Figure 2.

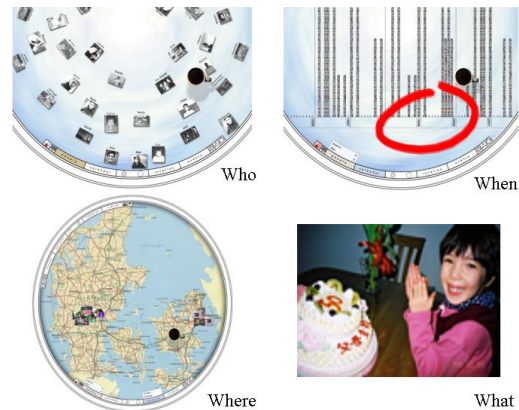


Figure 2. 4W(Who, When, Where, What)

* The figures are best viewed in color and at higher resolution at <http://www.ifp.uiuc.edu/~qitian/cvpr01tk/>

Other semantic annotation includes keywords and captions, as well as automatically extracted low-level visual feature vectors (color, texture, structure). Contents will be annotated in MPEG-7 format. (3) Content-based image layout: Smart layout, and user-guided image layout, using feature weighting vector estimation in high-dimensional feature space, and then projecting the images from high-dimensional feature space to 2-dimensional displaying space by Principle Component Analysis (PCA), with post linear and/or nonlinear processing.

There are a number of ways to implement this design. The physical sharing device can be either a touch sensitive table surface, or a rear projected or top projected horizontal table surface with infrared cameras to track hand motions. A portable device with projection and sensing capabilities can also be created so that the virtual table-top can travel with people to places. Our first prototype is constructed as a top projection table with a whiteboard as the table surface. This implementation provides a high resolution as well as a sturdy surface, which serves well as both a display device and a table to set familiar objects such as teacups on. We will add vision capabilities to the current table set up to enable both object avoidance in image display and multiple simultaneous control.

When people use the system they will cause different sets of their pictures to be displayed on the table. The pictures will be displayed in a radial-pattern in the story-space, so that pictures initially will be oriented towards the outer rim of the table. Users have the option of moving or re-orienting pictures using the touch-screen, in a rough simulation of how real pictures would slide along a table. By dragging one's hand along the outer rim of the table, the users can also spin the entire contents of the table, much like a lazy-susan.

The primary method of navigation is organized about four questions essential to storytelling: who?, when?, where?, and what?¹ Control panels located on the perimeter of the table contain buttons labeled "people", "calendar", "location", and "events", corresponding to these four questions. When a user presses the "location" button, for example, the display on the table changes to show a map of the world. Every picture in the database that is annotated with a location will appear as a tiny thumbnail at its location. The user can pan and zoom in on the map to a region of interest, which increase the size of the thumbnails. Similarly, by pressing one of the other three buttons, the user can cause the pictures to be organized by the time they were taken along a linear straight timeline, the people they contain, or the event-keywords, which the pictures were annotated with.²

¹ We do not currently support "why?", which is also useful for storytelling.

² We assume the pictures are *partially* annotated.

The user can form implicit Boolean queries, or filters, by selecting items or regions in the different navigational views. For example, if someone selects two friends in the people view, then (until this selection is retracted) only pictures containing one or both of these friends will be highlighted in subsequent navigation. If the user next selected "location", for example, then they would see where they have traveled with either of these friends by observing where the highlighted pictures appear on the map. If the user selected "calendar" instead, they would see when they have taken pictures with these friends. Another non-menu driven query metaphor used in PDH is invoked when the user presses and holds down a particular picture. The system then offers the user the ability to display pictures taken at a similar time, or a similar place, or with the same people, or at the same event as the selected picture.

At any point, the users can add or remove pictures from an initially blank "story-space". Users can add pictures based on their partial annotations or based on their similarity to other pictures, and can add them one at a time or in large groups. For example, users can use the above method of navigation to select all pictures that include a specified person over a specified time period and then press the "show" button to display all selected pictures in their story-space. In the following sections, we discuss how visualization and layout algorithms can help decide where to place images in the story-space, and can support the user's re-arrangement of pictures within this space.

3. Visualization and Layout Optimization

Traditional image database retrieval systems display query results as a list, sorted by similarity to the query. This presents one major drawback: relevant images can appear at separate places in the ordered list. The purpose of our proposed content-based visualization is augmenting a user's perception so as to visualize a large information space that cannot be easily perceived by traditional sequential array. The retrieved images are displayed not only in ranked order of similarity from the query but also according to their mutual pair-wise similarities, so that similar images are grouped together.

In our experiments, the 37 visual features (9 color moments [1], 10 wavelet moments [2] and 18 water-filling features [3]) are pre-extracted from the image database and stored off-line. The 37 visual features can be formed into a single feature vector and projected to the 2-D screen based on Principle Component Analysis (PCA). PCA is a very fast linear transformation that achieves the maximum distance preservation from the original high dimensional feature space to 2-D screen among all the linear transformations [4]. This visual layout is denoted as a "PCA Splat."

Figure 3 shows the top 20 retrieved images given the top left image as the query. Figure 4 shows an example of PCA Splats for the top 20 retrieved images shown in Figure 3. The sizes of the images are determined by their visual similarity to the query. The higher the rank, the larger the size. There is a number close to each image indicating its corresponding rank in Figure 3.

Clearly the relevant images are close to each other in the PCA Splats while these images are not consecutive to each other in the 1-D display in order of the decreasing similarities in Figure 3. PCA Splats also convey mutual distance information about all pair-wise similarities between images while the ranked 1-D display in Fig. 3 does not provide this perceptual information.

The layout is further optimized for maximal visibility by adjusting the size and position of each image to minimize overlap while relating other dimensions of information (such as relevance/similarity to the query). For details, interested readers refer to [5].

Figure 5 shows the optimized PCA Splats for Fig. 4. Clearly, the overlapping is minimized while the relevant images are still close to each other to allow a global view. With such a display, the user can see the relations between the images, better understand how the query performed, and subsequently express future queries more naturally.

4. Context and User Modeling

Image content and “meaning” is ultimately based on semantics. The user’s notion of content is a high-level concept, which is quite often removed by many layers of abstraction from simple low-level visual features. Even near-exhaustive semantic (keyword) annotations can never fully capture context-dependent notions of content. The same image can “mean” a number of different things depending on the particular circumstance.

By user-modeling or “context awareness” we mean that our system must be constantly aware of and adapting to the changing concepts and preferences of the user. A typical example of this human-computer synergy is having the system learn from a user-generated layout in order to visualize new examples based on identified relevant/irrelevant features. In other words, design smart browsers that “mimic” the user, and over-time, adapt to their style or preference for browsing and query display. Given information from the layout, *e.g.*, positions and mutual distances between images, a novel feature weight estimation scheme, noted as α -estimation is proposed, based on a non-negative least squares optimal solutions, where α is a weighting vector for feature *e.g.*, color, texture and structure (and semantic keywords). The algorithm will mimic the original (user) layout using the learned feature weights.

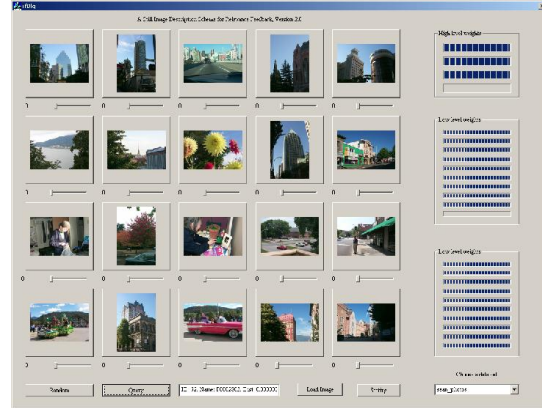


Figure 3. Top 20 retrieved images (ranked top to bottom and left to right; query image is shown first in the list)

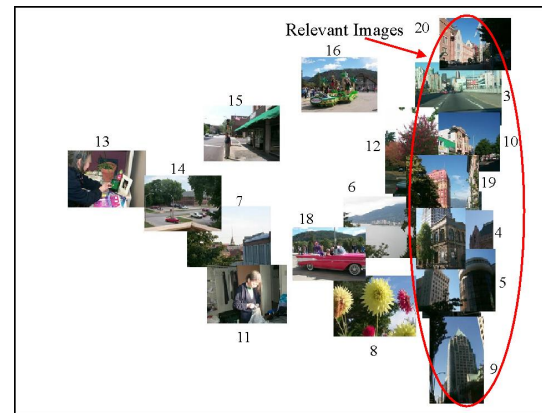


Figure 4. PCA Splat

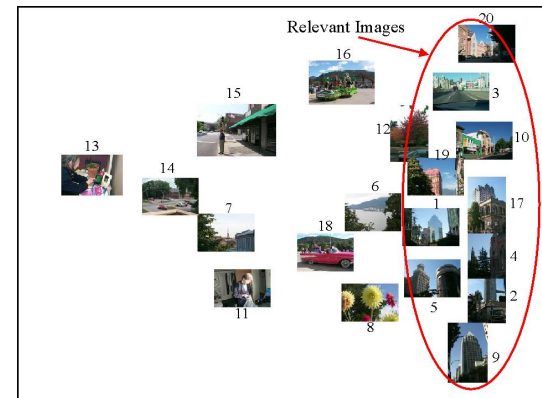


Figure 5. Optimized PCA Splat

Figure 6(a) shows an example of a layout. It is a PCA Splats of the images with their high dimensional features weighted by some (unknown) α . Figure 6(b) shows the corresponding computer-generated layout. It is also a PCA Splats of the images but with their high dimensional feature vectors weighted by the estimated α , which is determined solely from the configuration of Fig. 6(a). For comparison, Figure 6(c) shows the PCA Splat of the same images with their high dimensional feature vectors weighted by some random number (arbitrary α).

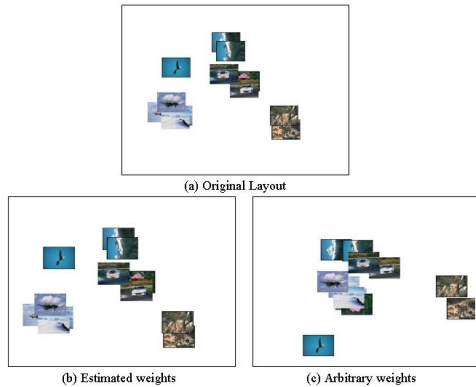


Figure 6. (a) An example layout

Computer-generated layout based on (b) reconstruction using learned feature weights, and (c) the control (arbitrary weights)

Given the lack of sufficiently large (and willing) human subjects, we undertook a Monte Carlo approach to testing our user-modeling and estimation method. Thereby simulating 1000 computer generated layouts (representing ground-truth values of α 's), which were meant to emulate 1000 actual user-layouts or preferences. In each case, α -estimation was performed to recover the original values as best as possible. As a control, 1000 randomly generated feature weights were used to see how well they could match the user layouts (*i.e.*, by chance alone).

The scatter-plot of α -estimation is shown in Figure 7. Clearly there is a direct linear relationship between the original weights and the estimated weights. In terms of actual measures of stress and deviation we found that the α -estimation scheme yielded the smaller deviation 78.4% of the time and smaller stress 72.9%.

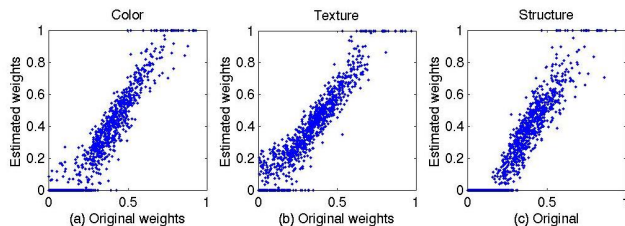


Figure 7. Scatter-plot of α -estimation: Estimated weights vs. original weights

We have also conducted a preliminary user study that has shown the superior performance of the proposed α -estimation over the random weighting as a control (or sanity check). There are two tests. The first one is to test whether the original feature weighting will generate a better layout than the control (*i.e.*, random weightings), which is denoted as " α -matters." The second is to test whether the estimated feature weights will generate a better layout than random weightings, and is denoted as α -estimation matters. 12 naïve users were instructed in

the basic rules for the user test and asked to perform either the α -matters and α -estimation-matters tests. The users were given the following instructions: (1) Both absolute and relative positions of images matter (2) In general, similar images, like cars, tigers, should cluster. (3) The relative positions of clusters matter. Each user performed a single forced-choice test only.

The average percentage of the vote rate for the original weight in α -matters test is 84.6% and user's consistency rate is 91.7%. The average percentage of the vote rate for the estimated weights in α -estimation-matters test is 95.8% and the user's consistency rate is 96.7% - indicated that the estimated weights are indeed favored over the random weights. (50% would indicate no difference or preference for estimation vs. random).

5. Discussions

The PDH project is at its initial stage. We have just begun our work in both the user interface design and photo visualization and layout algorithms. Many interesting questions still remain as our future research in the area of content-based information visualization and retrieval. The next task is to carry out an extended user-modeling study by having our system learn the feature weights from various sample layouts provided by the user. We have already developed a framework to incorporate visual features with semantic labels for both retrieval and layout. Incorporation of relevance feedback in our framework seems very intuitive and is currently being explored. Another challenging area is automatic "summarization" and display of large image collections. Since summarization is implicitly defined by user preference, α -estimation for user-modeling will play a key role in this and other high-level tasks where context is defined by the user.

References

1. M. Stricker and M. Orengo, "Similarity of Color Images", *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1995.
2. J. R. Smith and S. F. Chang, "Transform Features for Texture Classification and Discrimination in Large Image Database", *Proc. IEEE Intl. Conf. on Image Proc.*, 1994.
3. S. X. Zhou, Y. Rui and T. S. Huang, "Water-filling algorithm: A novel way for image feature extraction based on edge maps", in *Proc. IEEE Intl. Conf. On Image Proc.*, Japan, 1999.
4. Jolliffe, I.T., *Principal Component Analysis*, Springer-Verlag, New-York, 1986.
5. Q. Tian, B. Moghaddam, T. S. Huang, "Display Optimization for Image Browsing", will appear in *2nd Int'l Workshop on Multimedia Databases and Image Communications (MDIC'01)*, Sep. 17-18, 2001, Amalfi, Italy.